DOCUMENT RESUME

ED 177 166 TM 008 500

TITLE	APGA Policy Statement: Responsibilities of Users of Standardized Tests.
INSTITUTION	American Personnel and Guidance Asscciatien, Washington, D.C.
PUB DATE	5 Oct 78
NOTE	8p.
JOURNAL CIT	Guidepost; p5-8 Oct 5 1978
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Codes of Ethics; Counselors; Examiners; Information Utilization; Policy Formation; *Psychclcgical Testing; Qualifications; Scoring; *Standardized Tests; Standards; Testing; *Test Interpretation; *Test Results; *Test Selection
IDENTIFIERS	*American Personnel and Guidance Association

ABSTRACT

This statement presents the pcsition cf and addresses the needs of the professional members of the American Personnel and Guidance Association (APGA) divisions and regions. The seven sections discuss: (1) the development of the statement; (2) guidelines for avoiding charges of irresponsible practice; (3) qualifications of test users (those who select, administer, score, or interpret tests); (4) test selection; (5) test administration; (6) test interpretation--including scores, norms, reliability, and validity; and (7) communication of test results. (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

*APGA Policy Statement: Responsibilities of Users of Standardized Tests, from Guidepost, October 5, 1978, pp. 5-8.

* Copyright 1978, American Personnel and Guidance Association. Reprinted with permission.

ED177166

TM008 500

INTRODUCTION

During the past several years, individual APGA members have been under increasing pressure from their various constituencies to define, provide and employ safeguards against the misuse of standardized tests. APGA as an organization has also been challenged by individuals and agencies to provide leadership in the face of growing concern about the effects of testing on clients of all ages and all subpopulations, and in all settings.

At the 1976 APGA convention, the board of directors requested action on the development of a statement on the responsible use of standardized tests. A committee representing all APGA divisions and regions spent two years studying the issues and developing the following statement.

Target audience: The statement is intended to present the position and address the needs of the professional members of APGA divisions and regions. Although this position may provide guidance for test developers, teachers, administrators, parents, press or the general public, it is not designed to represent these audiences. The statement is built on the assumption that test data of themselves are neutral and that guidelines are needed to promote constructive use of tests.

Organization and focus: The statement is organized into eight sections: Introduction, Decision Rules, Test Selection, Qualifications of Test Users, Test Administration, Scoring of Tests, Test Interpretation and Communication. Each section is directed toward the various uses and decisions that must be made by the test user (e.g., whether to test, which test(s) to use, what data to obtain, how to interpret, etc.). The committee developed a classification system for the uses of standardized tests and treated only those issues that fit into the use classification scheme.

The next step was to define issues related to the classification system. Issues were sought from individuals, professional statements, literature and the popular press. Issues were examined in terms of their relevance to APGA members and to their importance in terms of the possible consequences to the person(s) tested.

Only the principles underlying each issue are specified. These principles are appropriate as standards for all APGA divisions and regions. Divisions and regions are encouraged to develop their own statements, expanding on each prrinciple with specific procedures and examples and appropriate to their members. The principles are grouped around similar issues and are indexed for easy reference.

Composition of Committee: Each division president and the representative of each region was asked to appoint to the committee a member who was: (1) knowledgable in the use of standardized tests; (2) aware of the national concerns about the use of tests; (3) willing to involve the division/region membership in identification of the concerns, needs and propositions of the division/region as these relate to the responsible use of tests; (4) able to participate actively and to respond promptly to requests for review of draft documents; (5) capable of securing cooperation of division/region officers in developing procedures for division/region implementation of the statement; (6) willing to accept specific assignments such as providing current organizational statements.

In addition to the division/region appointment, the committee included two members-at-large and a chairperson, all appointed by the APGA president, and three test experts to serve as a core committee to assist the chairperson in analyzing and synthesizing input from committee members and in preparing draft statements for committee review.

To furnish perspective for the work of this committee, a review of relevant literature was conducted and each member of the committee received copies of numerous position papers, reports, articles and monograps that added to understanding the issues and the consequences of alternative principles.

Among these papers were the interim report of the 1975-76 APGA Committee on Standardized Testing; the 1972 APGA/AMEG Statement on the Responsible Use of Tests; position papers of individual divisions and of other professional organizations such as the American Psychological Association, the American Education Research Association and the National Council on Measurement in Education; journal articles; and conference presentations.

The committee's statement is intended to be sensitive to current and emerging problems and concerns that are generic to all APGA divisions/regions and to address these problems and concerns with principles that are specific enough to serve as a template to develop division/region statements addressed to the specific disciplines/settings of individual divisions/regions.

DECISION RULES

In human service agencies, decisions about client needs may be made on the basis of direct observation or historical information alone. Further refinement of direct observation and historical data can often be obtained by employing standardized tests.

Deciding whether to test creates the possibility of three classes of errors relative to the agency functions of description, diagnosis, prescription, selection, placement, prediction, growth evaluation, etc.

First, a decision not to test can result in misjudgments that stem solely from inadequate data.

Second, tests may be used well, producing data that could improve accuracy in decisions affecting the client but that are not utilized.

Third, tests may be misused through inappropriate selection, improper administration, inaccurate scoring, incompetent interpretation or indiscriminate, inadequate, or inaccurate communication.

To reduce the chance for errors, the responsible practitioner will always determine in advance why a given test should be used. This provides protection and benefits for both the client and the agency. Having a clearly developed rationale increases the probable benefits of testing by indicating how a particular set of information, when used by an individual or set of individ-

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY APGA TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

uals, will contribute to a sounder decision without prejudice to either the client or the agency.

The guidelines that follow are intended to provide decision rules to help agencies and practitioners avoid charge of irresponsible practice.

Defining purposes for testing:

1. Decide whether you will be testing to evaluate individuals, groups or both.

2. Identify your interests in the particular target population in terms of the agency's purposes and capabilities.

3. Determine limits to diagnosis, prediction or selection created by age, racial, sexual, ethnic or cultural characteristics of those to be tested.

4. Develop specific objectives and limits for the use of test data in relation to each of the component service areas of placement/selection, prediction (expectancies), description/diagnosis and growth studies (assessing change over time).

A. Placement: If the purpose is selection or placement (selection is a simple in-out sort of placement), the test selector and interpreter must know about the programs or institutions in which the client may be placed and be able to judge the consequences of such placement or exclusion for the client.

B. Prediction/expectancies: If the purpose is prediction, the persons deciding to test and/or interr ret the results must understand the pitfalls of labeling, stereotyping and prejudging people. Ways to avoid these potentially invidious outcomes should be known.

C. Description/diagnosis: If the purpose is diagnosis or description, the selector or interpreter should understand enough about the general domain being measured to be able to identify those aspects adequately measured and those not.

D. Growth/change assessment: If the purpose is to examine growth, the person designing the study and interpreting the results needs to know the many problems associated with such measurement:

1) the unreliability of change measures;

2) the pitfalls in using norms as reference points;

3) the associated problems of articulation and comparability;

4) the limitations of scoring scales, such as grade equivalents, that may not have the comparable meaning which they appear to have at different scale levels.

Determining information needs:

1. Assess the consequences for the clients of both testing or not testing.

2. Determine what decisions can be made with existing information to avoid unnecessary data-gathering efforts.

3. Limit data gathering to those functions or aptitude, achievement, interests/attitudes/values and perceptual-motor skills that are directly relevant in making decisions about delivery of services to a particular individual or group.

4. Identify whether the test being considered can provide acceptable levels of precision (reliability) for the decision being made.

5. Identify whether the data obtained can be cross-validated against other available data as a part of the decision-making process.

6. Determine the amount and form of data to be shared on the basis of maximum relevance to the agency's purposes and capabilities.

Identifying users of test information:

1. Data should be prepared so that they can be comprehended by the persons using the data for decision making.

2. Limit access to users specifically authorized by the law or by the client.

3. Identify obsolescence schedules so that stored personal test data may be systematically reclassified and relocated to historical files or destroyed.

4. Process personal data used for research or program evaluation so as to assure individual anonymity.

QUALIFICATIONS OF TEST USERS

All professional personnel and guidance workers should have formal training in psychological and educational measurement and testing. Nevertheless, it is unreasonable to expect that this training necessarily makes one an expert or even that an expert always has all the knowledge and skill appropriate to any particular situation. Thus, questions of user qualifications should always arise when testing is being considered.

Those who participate in any aspect of testing should be qualified to do so. Lack of proper qualifications leads to misuse, errors and sometimes damage to clients. Each professional is responsible for making judgment on this matter in each situation and cannot leave that responsibility either to clients or to those in authority.

In many instances information or skills that may be lacking can be acquired quite readily by those with a background of professional training and experience. In all instances it is incumbent upon the individual to obtain that training or arrange for proper supervision and assistance when engaged in, or planning to engage in, testing.

The requisite qualifications for test users depend on four factors: (1) the particular role of the user: (2) the setting in which the use takes place; (3) the nature of the test; and (4) the purpose of the testing.

These factors interact with each other but may nevertheless be considered separately for the purposes of these standards.

Roles of test users, selectors, administrators, scorers and interpreters:

A test user may play all of these roles or any subset of them when working with other professional personnel. In some situations each role may be the responsibility of a different person. The knowledge and skills that pertain to these roles are listed under the sections so headed. The general principle is that the test users should engage in only those testing activities for which their training and experience qualify them.

Settings and conditions of test use:

Counselors and personnel workers should assess the quality and relevance of their knowledge and skills to the situation before deciding to test or to participate in a testing program.

Characteristics of tests:

Tests differ in many ways, and users need to understand the peculiarities of the instruments they are using.

Purposes of testing:

The purpose of the testing dictates how the test is used and thus may influence requisite qualifications of users beyond those entailed by their testing roles. Technically proper use for ill-understood purposes may constitute misuse.

TEST SELECTION

Tests should be selected for a specific measurement purpose, use and interpretation. The selection of tests should be guided by information obtained from a careful analysis of the following major considerations:

— What are the characteristics of the population to be tested?

—What knowledge, skills, abilities or attitudes are to be assessed?

— What are the purposes for testing?

— How will the test scores be used and interpreted?

When complete answers to these questions have been obtained, selection or development of tests should be directed toward obtaining measures that are congruent with the stated needs for assessment in terms of the purposes, content, use, interpretation and particular characteristics of the individuals who are to be tested.

Selection of tests must also be guided by the criteria of technical quality recommended by the measurement profession and published by APA/AERA/NCME in "Standards for Educational and Psychological Tests" (1974). Full recognition and analysis of these considerations should become the focus of a process to select appropriate tests. The responsible test selector will:

Select appropriate tests:

1. Select tests that have been demonstrated, to the satisfaction of professional specialists, as appropriate for the characteristics of the population to be tested.

2. Select tests that are within the level of skills of administration and interpretation possessed by the practitioner.

3. Determine whether a common test or different tests are required for the accurate measurement of groups with different characteristics.

4. Recognize that different tests for cultural, ethnic and racial groups constitute inefficient means for making corrections for differences in prior life experiences, except where different languages are involved.

5. Determine whether persons or groups that use different languages should be tested in either or both languages and in some instance by prior testing for bilingualism.

Relate evidence or validity to particular usage:

1. Apply tests or selection only when they show predictive validity for the specific tasks or competencies needed in an educational or employment assignment to maintain legal prescriptions for non-discriminatory practices in selection, employment or placement.

3. Determine validity of a test (whether the test measures what it claims to measure) through evidence of the contructs used in developing the measures, the correlation of the test performance with another appraisal of the characteristics being measured, or the predictions of specified behavior from the test performance.

3. Determine that the content of the test has high congruence with the users' definition of the knowledge and skills that are the desired criteria of human performance to be appraised.

4. Confirm that the criteria of human performance to be appraised are contained in the tasks and results of the testing procedure.

Employ user participation in test selection:

Actively involve the persons who will be using the tests (administering, scoring, summarizing, interpreting, making decisions) in the selection of tests that are congruent with the locally determined purposes, conditions and uses of the measurement.

Select tests to satisfy local use:

1 Give specific attention to how the test is designed to handle the variation of motivation among persons taking the test, the variation or bias in response to the test content and the effects of the presence or absence of guessing in the responses to the test questions.

2. Determine whether tests standardized for nationwide use show evidence that such tests yield comparable results for individuals or groups with cultural differences.

3. Identify and analyze the effects of working speed and language facility in relation to the criteria of human performance that are expected to result from the test.

Consider technical characteristics of tests:

1. Select only published or locally developed tests that have documented evidence of the reliability or consistency of the measure.

2. Select tests that have documented evidence of the effectiveness of the measure for the purpose to be served: placement/selection, prediction (expectancy), description/diagnosis, or growth studies (change over time). A test is rarely equally effective for the four common test uses.

3. Consider the procedures used in standardization and norming for relevance to the local population and the desired use and interpretation.

4. Use separate norms for men and women only when empirical evidence indicates this is necessary to minimize bias.

5. Determine the degree of reliability (or validity) demanded of a test on the basis of the nature of the decisions to be based on test scores.

6. A test for final diagnosis or selection requires a higher degree of reliability than an initial screening test.

7. Explicitly list and use the ease and accuracy of the procedures for scoring, summarizing and communicating test performance as criteria for selecting a test.

8. Recognize that the technical characteristics and norms of standardized tests may vary when used with different populations. The selection process should include trial administrations to verify that the test is functioning with the technical characteristics and desired results for the local population and local uses.

Practical constraints of cost, conditions and time for testing must be considered but not used as the primary criteria for test selection.

TEST ADMINISTRATION

Test administration includes all procedures that are used to ensure that the test is presented consistently in the manner specified by the test developers and used in the standardization and that the individuals being tested have orientation and conditions that maximize opportunity for optimum performance.

Standardized tests should provide manuals giving specific directions for administering, scoring and interpreting tests. Tests developed for a specific local purpose, use or population should be administered in a prescribed and consistent manner to obtain optimum

performance from the individuals being tested. Effective administration of tests requires that the administrator have knowledge and training with the instruments and the processes of presentation.

Orientation:

1. Inform testing candidates, relevant institutions or agencies and the community about the testing procedure. The orientation should describe the purposes and contents sampled by the test, how it is administered and how the scores will be reported and used.

2. Provide annual training for test administrators by qualified professional specialists if your agency or institution uses tests or sponsors testing programs.

3. Routinely review the test materials and administration conditions well in advance of the time for testing so that full preparation will ensure standardized administration and recognition of any irregularities that may occur.

4. Ensure that the orientation is sufficient to make the test relevant for the individual or group being tested before beginning test administration.

5. Ensure that all persons being tested have the specified practice with sample problems or test taking skills prior to their performance on the test.

6. Demonstrate the techniques and requirements for marking machine-scorable answer sheets. Check all individuals taking a test for competency in the techniques of recording their answers prior to the specific period of testing.

Qualifications of test administrators:

1. Administer standardized tests only if you are qualified by training and experience as competent to administer particular tests.

2. Know the exact population and procedures used in standardizing the test and determine that the test is appropriate for the local population that is to be tested.

3. Acquire extensive training required to administer, score or interpret tests requiring test-specific training.

Giving directions:

1. Administer standardized tests with the verbatim instructions, exact sequence and timing and the identical materials that were used in the test standardization.

2. Present all tests (whether standardized, published or locally constructed) in an identical manner to ensure that the test is a fair and comparable demonstration of the performance of each individual taking the test.

Recognize that taking a test may be a new and frightening experience or stimulate anxiety or frustration for some individuals. Communicate to the examinees that they should attempt each task with positive application of their skills and knowledge and the anticipation that they will do their best.

Testing conditions:

1. Devote concentrated attention to observing the condition and reactions of the individuals being tested. Observe those being tested and identify environmental, health or emotional conditions that should be recorded and considered as invalidating elements for the test performance.

2. Possess and demonstrate clear verbal articulation, calmness and positive anticipation, empathy for and social identification with the examiness, and, impartial treatment for all being tested.

3. Determine whether the testing environment (seating, work surfaces, lighting, heating, freedom from distractions, etc.) is conducive to the best possible performance of the test-takers.

4. Administer tests in physical facilities and psychological climates that allow each individual being tested to achieve optimum performance.

5. Record any deviation from standardized test administration procedure (such as used to accommodate handicapping conditions) and make it a permanent attachment to the test score or record.

6. Develop and complete for each test a systematic and objective procedure for observing and recording the behavior of those being tested (and recording conventional or deviant conditions of testing). Attach this record to the test scores of the persons tested.

7. Provide a written record of any circumstances that may have increased or reduced the opportunity of an individual being tested to demonstrate his or her best performance.

Accept responsibility for seeing that invalid or questionable test scores are not recorded, or not recorded without written qualification of the conditions that may have affected optimum test performance.

9. Arrange assistance from trained personnel in providing uniform conditions and in observing the conduct of the examinees when large groups of individuals must be tested.

Professional collaboration:

Recognize that in institutional settings, and wherever skill and knowledge can be pooled and responsibility shared, it is the qualifications of the team as a whole that count rather than those of individuals. However, coordination and consistency must be maintained.

TEST SCORING

The measurement of human performance depends on accurate and consistent application of defined procedures for crediting the responses made by persons being tested. The procedures for scoring and recording test performance must be continuously audited for consistency and accuracy.

1. Routinely rescore a sample of the test answer sheets to verify the accuracy of the initial scoring.

2. Employ systematic procedures to verify the accuracy and consistency of machine scoring of answer sheets.

3. Obtain a separate and independent verification that appropriate scoring rules and normative conversions are used for each person tested.

4. Verify as accurate the computation of raw scores and the conversion of raw scores to normative or descriptive scales prior to release of such information to the tested person or to users of the test results.

5. Routinely check machine or manual reports of test results for accuracy. The person performing this task must be qualified to recognize inappropriate or impossible scores.

6. Develop and use systematic and objective procedures for observing and recording the conditions and behaviors of persons being tested and make this a part of the scores or test results that are reported.

7. Clearly label the scores that are reported and the date that a particular test was administered.

TEST INTERPRETATION

Test interpretation encompasses all the ways we assign value to the scores.

A test can be described as a systematic set or series of standard observations of performances that all fall in some particular domain. Typically each observation yields a rating of the performance (such as right or wrong and pass or fail), then these ratings are counted and this count becomes the basis of the scores. Such scores are usually much more stable than the result of any single performance. This score reliability creates the possibility of validity greater than can be obtained from unsystematic or nonaggregated observations.

The proper interpretation of test scores starts with understanding these fundamental characteristics of tests. Given this, the interpretation of scores from a test entails knowledge about (1) administration and scoring procedures; (2) scores, norms, and related technical features; (3) reliability; and (4) validity.

Adequate test interpretation requires knowledge and skill in each of these areas. Some of this information can be mastered only by studying the manual and other materials of the test; no one should undertake the interpretation of scores on any test without such study.

Administration and Scoring:

Standard procedures for administering and scoring the test limit the possible meanings of scores. Departures from standard conditions and procedures modify and often invalidate the criteria for score interpretation.

1. The principles in the section on administration and scoring need to be understood by those engaged in interpretation.

2. Ascertain the circumstances peculiar to the particular administration and scoring of the test.

A. Examine all reports from administrators, proctors and scorers concerning irregularities or conditions, such as excessive anxiety, which may have affected performance.

B. Weigh the possible effects on test scores of examiner-examinee differences in ethnic and cultural background, attitudes and values in light of research on these matters. Recognize that such effects are probably larger in individual testing situations.

c. Look for administrators' reports of examinee behavior that indicate the responses were made on some basis other than that intended — as when a student being tested for knowledge of addition-number-facts adds by making tallies and then counting them.

3. Consider differences among clients in their reaction to instructions about guessing and scoring.

4. Recognize or judge the effect of scorer biases and judgment when subjective elements enter into scoring.

Scores, norms and related technical features:

The result of scoring a test is usually a number (or a set of numbers) called a raw score. Raw scores taken by themselves are not usually interpretable. Some additional steps must be taken.

The procedures either translate the numbers directly into descriptions of their meaning (e.g., pass or fail) or into other numbers called derived scores (e.g., standard scores) whose meaning stems from the test norms.

To interpret test scores, these procedures and the resulting descriptions or derived scores need to be thoroughly understood. Anything less than full understanding is likely to produce at least some, and probably many, serious errors in interpretation. The following are imperatives for interpreting tests:

1. Examine the test manuals, handbooks, users' guides and technical reports to determine what descriptions or derived scores are produced and what unique characteristics each may have.

2. Recognize that direct score interpretations such as mastery and nonmastery in criterion-referenced tests depend on arbitrary rules or standards.

A. Report number or percent of items right in addition to the indicated interpretation whenever it will help others understand the quality of the examinee's test performance.

B. Recognize that the difficulty of a fixed standard, such as 80 percent right; will vary widely from objective objective. Such scores are not comparable in the normative sense.

C. Recognize that when each score is classified as pass fail, mastery-nonmastery or the like, that each element is being given equal weight.

3. Use the derived scores that fit the needs of the current use of the test.

A. Use percentile ranks for direct comparison of individuals to the norm or reference group.

B. Use standard scores or equal unit scaled scores whenever means and variances are calculated or other arithmetic operations are being used.

4. Recognize that only those derived scores that are based on the same norm group can be compared.

5. Consider the effect of any differences between the tests in what they measure when one test or form is equated with another, as well as the errors stemming from the equating itself.

Give greater credence to growth or change shown by the same test (including level and form) than to equated measures except where practice effects or feedback have destroyed the validity of a second use.

6. Evaluate the appropriateness of the norm groups available as bases for interpreting the scores of clients.

A. Use the norms for the group to which the client belongs.

B. Consider using local norms and derived scores based on these local norms whenever possible.

7. Acquire knowledge of specific psychological or educational concepts and theories before interpreting the scores of tests based on such knowledge.

Reliability:

Reliability is a prerequisite to validity. Generally, the greater the number of items the greater the reliability of the test. The degree to which a score or a set of scores may vary because of measurement error is a central factor in interpretation.

1. Use the standard error of measurement to obtain a rough estimate of the probable variation in scores due to unreliability.

2. Use the reliability coefficient to estimate the proportion of score variance that is not due to error.

3. Consider the sources of variance attributable to error in the particular reliability indexes reported in relationship to the uses being made of the scores.

4. Assess reported reliabilities in light of the many extraneous factors that may have artificially raised or lowered these estimates, such as test speededness, sample homogeneity or heterogeneity, restrictions in range and the like.

5. Distinguish indexes of rater reliability (i.e., of objectivity) from test reliability.

Validity:

Proper test interpretation requires knowledge of the validity evidence available for the test as used. Its validity for other uses is not relevant. The purpose of testing dictates how a test is used. Technically proper use for ill-understood purposes may constitute misuse. The nature of the validity evidence required for a test is a function of its use.

Prediction — developing expectancies: The relationship of the test scores to an independently developed criterion measure is the basis for predictive validity.

1. Consider both the reliability and the relevance of the criterion measures used.

2. Use cross validation data to judge the validity of predictions.

3. Question the meaning of an apparently valid predictor that lacks both construct and content validity. Assess the role of underlying and concomitant variables.

4. Consider the validity of a given measure in the context of all the predictors used or available. Does the measure make an independent contribution to the prediction over and above that provided by other measures?

5. Consider the pitfalls of labeling, stereotyping and prejudging people. The self-fulfilling prophecies that may result are often undesirable.

Placement/selection: Predictive validity is the usual basis for valid placement. Consider the evidence of validity for each alternative (i.e., each placement) when inferring the meaning of scores.

1. Obtain adequate information about the programs or institutions in which the client may be placed in order to judge the consequences of such placement.

2. Estimate the probability of favorable outcomes for each possible placement (e.g., both selection and rejection) before judging the import of the scores.

3. Consider the possibility that outcomes favorable from an institutional point of view may differ from those that are favorable from the examinees' point of view.

4. Examine the possibility that the clients' group membership (race, sex, etc.) may alter the reported validity relationships.

5. Use all the available evidence about the individual to infer the validity of the score for that individual. Each single piece of information about an individual, (e.g, test score, teacher report, or counselor opinion) improves the probability that proper judgments and decisions can be made.

A. Test scores should be considered in context; they do not have absolute meaning.

B. Single test scores should not be the sole basis for placement or selection.

Description/diagnosis: Distinguish between those descriptions and diagnoses using psychological constructs that can be validated only indirectly and those for which content specifications suffice.

1. Identify clearly the domain specified by those asserting content validity. Assess the adequacy of the content sampling procedures used in writing and selecting items.

2. Identify the dimensions of the construct being measured when multiple scores from a battery or inventory are used for description.

A. Examine the content validity and/or the construct validity of each score separately.

B. Consider the relative importance of the various subtests, parts, objectives or elements yielding separate scores and judge the weight they each should be given in interpretation.

C. Recognize that when scores are summed or averaged their weight is a function of their variances.

D. Recognize that when each score is classified as pass-fail, mastery-nonmastery or the like, each element is being given equal weight.

3. Examine the completeness of the description provided, recognizing that no set of test scores completely describes a human being.

Growth — Studies of change: Valid assessment of growth or change requires both a test having descriptive validity and a procedure for establishing that the scores obtained differ from those that might arise when no change has occurred.

1. Report as possibilities all the interpretations the study or evaluation design permits. Point out those interpretations that are precluded by the design used.

A. When standard procedures such as the RMC models (see Tallmadge, GK. and Horst, D.P., A procedural guide for validating achievement gains in educational projects; Mountain View, Calif: RMC Research Corp., Dec. 1975) for measuring growth in achievement are employed, use the descriptions of strengths and weaknesses provided.

B. Look for naturally occurring control groups not part of design whenever possible.

2. Consider the strengths and weaknesses of the particular tests used with respect to this use.

Consider the possibility of floor or ceiling effects, the content changes level to level, the adequacy of articulation in multilevel tests, the comparability of alternate forms, the adequacy of the score-equating across forms, and the comparability of timing of the testing to that of the norming.

3. Recognize the unreliability of individual score differences as measures of change.

4. Recognize the limitations of scoring scales such as grade equivalents that may not have the comparable meaning they appear to have at different levels of the scale.

5. Recognize the need for equal interval scales when trying to assess the amount of change.

COMMUNICATING TEST RESULTS

Communication consists of reporting data in such a way that it is comprehensible and informative. The responsible practitioner reports test data with a concern for the user's need for information and the purposes of evaluating the significance of the information.

There must also be a concern for the right of the individual tested to be informed regarding how the results will be used for his or her benefit (informed consent), who will have access to the results (right to

privacy), and what safeguards exist to prevent misuse.

Where standardized test data are being used to enhance decisions about an individual, the practitioner's responsibilities are as follows:

Know the manual:

1. Become thoroughly familiar with the publisher's manual before attempting to "explain" any results.
2. Develop skills needed to communicate results of tests, using concepts that are frequently misunderstood before communicating results to clients, the public, or other recipients of the information.

Know the limits:

1. Inform the person receiving the test information that "scores" are approximations, not absolutes, and indicate the SEM or the margin of error in some other way, such as by reporting score intervals rather than points.
2. Candidly discuss with the person receiving the test information any qualifications necessary to understand potential sources of bias for a given set of test results relative to their use with a specific individual.
3. Emphasize that test data represent just one source of information and should rarely, if ever, be used alone for decision making.

Informed consent:

1. Inform the person receiving the test information of any circumstances that could have affected the validity or reliability of the results.
2. Inform the examinee of what action will be taken by the agency and who will be using the results.
3. Obtain the consent of the examinee before using test results for any purpose other than that advanced prior to testing.

Right to privacy:

Inform the examinee of steps to be taken to correct any erroneous information that may be on file as a result of testing.

Where standardized test data are being used to describe groups for the purpose of evaluation, the practitioner's responsibilities are as follows:

Background information:

1. Include background information to improve the accuracy of understanding about any numerical data.
2. Identify the purposes for which the reported data would be appropriate.

Politics:

Be aware that public release of test information provides data for all kinds of purposes and that some of these may be adverse to the interests of those tested.

Averages and norms:

1. Clarify in particular that "average" on a standardized test is a range, not a point, and typically includes the middle 50 percent of the group being considered.
2. Qualify all group data in terms of the appropriateness of the norms for that group.

Agency policies:

1. Work for agency test-reporting policies designed to strengthen and protect the benefits of the groups being measured.
2. Work within the agency to establish procedures for periodic review of internal test use.